学校编码：10384　　分类号____密级____

学号：　　UDC_____

硕 士 学 位 论 文

音频哼唱检索算法研究

Study on the Audio Query Algorithms by Humming

刘夙

指导教师姓名： 洪景新 高级工程师
专 业 名 称：信号与信息系统
论文提交日期：2012 年 月
论文答辩时间：2012 年 月
学位授予日期：2012 年 月

答辩委员会主席：_____

评 阅 人：_____

2012 年 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（　　）课题（组）的研究成果，获得（　　）课题（组）经费或实验室的资助，在（　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

摘要

近几年，随着网络上多媒体数据的增加，如何采用更加方便的方法进行快速检索已经成为一个研究的热点。特别是对于音频来讲，由于音乐数据的大量增加，一种能够更加有效检索这些海量数据的检索系统成为了目前的迫切需要。传统的检索方式是基于文本进行检索的，这需要对音乐文件进行文本信息的标注，例如歌名、演唱者、作曲者等。而基于内容的哼唱检索方式仅需要用户哼唱一小段旋律，就可以检索出相应的歌曲。这种新的检索方式不仅减小了音乐库所需的存储空间，同时也为用户检索音乐提供了极大的方便。

哼唱系统将采集到的用户哼唱信号经过预处理模块进行语音增强，在特征提取模块中对增强过的信号进行特征提取，用提取到的特征信息和音乐特征库中的特征信息进行旋律相似度计算，最后得出检索排名。

本文对哼唱检索（QBH）的主要模块——预处理模块、特征提取模块和旋律匹配模块进行了研究和改进。

首先，分析比较了几种语音增强算法的优缺点，采用了基于变参的改进减谱法作为哼唱检索的预处理模块。经过测试，变参的减谱法在噪声较大的情况下对于哼唱检索的排名提高有较好的效果，有效率为 93.3%。

其次，鉴于哼唱旋律音调高低和节奏快慢的不同，提出了对于提取到的语音特征进行归一化处理的方法。这种方法有效的提高了哼唱检索的排名，有效率为 85.7%。

最后，考察了几种旋律相似度计算的方法，构建了一个两层结构的旋律匹配模块。第一层采用 EMD 算法对旋律进行初步检索；第二层采用 DTW 算法对第一层检索排名前 20 的序列进行进一步检索，计算得到的 DTW 旋律相似度与 EMD 旋律相似度进行加权和，得到最终的检索排名。同时，本文针对传统 EMD 算法和 DTW 算法进行改进，提出基于权值均方的 EMD 算法和基于二维变量的 DTW 算法。经过测试，EMD 算法的改进有效率为 85.7%，DTW 算法的改进有效率为 79.2%。

关键词：哼唱检索，语音增强，旋律匹配

Abstract

In recent years, the method for querying multimedia data has become a research hotspot with the increasing data in network. Especially to audio, there is an urgent demand in building a more efficient query method to cope with the mass audio data. The traditional query method, which is based on text, need label music files, such as the name of music, the singer, the composer, etc. However，the query method named Query By Humming(QBH), which is based on content, only needs a small piece of melody sung by user, and then the corresponding music can be retrieved. The new query mothod decreases the storage space needed by music library and is more convenient to user.

In QBH, the humming signal is enhanced in reprocessing model. And then, the voice feature is extracted from enhanced signal in characteristic extraction model. Finally, the melody similarity is calculated using voice features of humming signal and music characteristic library and the query rank is obtained.

This paper suggests three aspects of improvement in three models of QBH respectively, which is preprocessing model, characteristic extraction model and melody matching model.

Firstly, through analyzing and comparing the advantages and disadvantages of several speech enhancement algorithms, the improved spectral subtraction based on varying parameter is chose in preprocessing model. The test result proves that the spectral subtraction based on varying parameter can effectively enhance the query ranking, and the effective rate is 93.3%.

Secondly, in consideration of the difference in pitch and rhythm, the normalization of the extracted voice features is proposed in characteristic extraction model. The query ranking is increased because of the normalization, and the effective rate is 85.7%.

Finally, this paper researches and analyzes several algorithms of computing

melody similarity, and then a two-step melody matching model is construct. In Step 1, the Earth Mover's Distance(EMD) algorithm is used to query the musics preliminarily; In Step 2, the Dynamic Time Warping(DTW) algorithm is used to query the musics which are ranked in the top 20 in Step 1. The weighted sum of melody similarities stemed from EMD and DTW is the final melody similarity, and then the query ranking is obtained according to the melody similarity. In the meanwhile，in contrary to traditional EMD and DTW, the improved EMD based on weight mean square and the improved DTW based on two dimensions is proposed in this paper. The test reault indicates that the effective rates of improved EMD and improved DTW are 85.7% and 79.2% respectively.

Key Words：Query By Humming, Speech Enhancement, Melody Maching

目 录

Contents